# Detection of Potentially Students Drop Out of College in Case of Missing Value Using C4.5

*by* Siti Mutrofin

# Detection of Potentially Students Drop Out of College in Case of Missing Value Using C4.5

[1]Siti Mutrofin, [5]Abdul Muiz Khalimi, [6]Eddy Kurniawan
*Information System*
*Universitas Pesantren Tinggi Darul Ulum*
Jombang, Indonesia
[1]sitimutrofin@ft.unipdu.ac.id

[2]Raden Venantius Hari Ginardi, [3]Chastine Fatichah
*Infromatic Engineering*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia

[4]Yuita Arum Sari
*Informatics Engineering*
*Brawijaya University*
Malang, Indonesia

***Abstract*—The reputation of a university can be determined by the number of students drop out. This problem was experienced by many universities in Indonesia. It has been conducted by many researchers, however the data acquisition, attributes were not well explained. This study is aiming for giving projection related to the reasons behind students drop out by using machine learning technique. The challenging phase of preprocessing primary datasets are missing value, balanced class distribution, and a variety of data types. Two classes are applied: drop out and graduate students. By analyzing the problem of missing value data, it can reflect the basis of why students drop out or students who have the potential to drop out. According to the problem of balanced class distribution, Decision Tree algorithm is utilized, meanwhile for tackling the various of data types, we use C4.5. The result shows that 20 attributes using stratified sampling is the best of among all datasets and experimentations with an average AUC, accuracy, precision, and recall values of 0.98, 96.87, 98.75, and 97.84 respectively. It indicates that the proposed method is suitable for predicting students drop out with a balanced case of class distribution, despite having a missing data value problem.**

***Keywords*—Balanced Class Distribution; C4.5; Drop out; Missing value.**

## I. Introduction

One common problem among universities in Indonesia are facing a growing number of students who drop out (DO students) per year, both in public and private universities. In Java Island, the highest number of DO students is in East Java Province [1]. The accumulation of DO students' numbers interprets the quality of the university [1, 2]. DO students which are usually associated with economic problems [3, 4, 5], bad motivation in learning [4, 5], and bad time management in college because they are also working. Some students move [6] to another university was because they accepted at a better university in the middle of their study or they have been accepted to work in another city. However, some students unintentionally expelled from the university because they have problems in academic or non-academic. DO students usually occurs in the first to the third year of their study, largely influenced by several factors: the economy, they got accepted as an employee, or accepted by another university that usually has better quality/reputation than the former university. On the other hand, many of the DO students are in the final year of study because they are unable to complete their thesis or found violating the university/academic rules. In this digital age, technology plays an important role to solve the problem in many ways, including this increasing number of DO students.

Industrial Revolution 4.0 and Society 5.0 are currently implemented. We are currently in the era of Society 5.0, where this era is more humanistic as proposed by the Japanese government in January 2016 [7]. Compared to The Fourth Industrial Revolution (4IR) proposed by German professor Klaus Martin Schwab, Society 5.0 aims to make people more qualified by utilizing technology and concentrating on the realization of 17 Sustainable Development Goals (SDG), established by the United Nations (UN) on September 25, 2015, in New York, USA [7]. The fourth objective of the SDG, "Quality Education", was proposed to make developing countries such as Indonesia overcome the problem of student oversight, to provide quality education for all Indonesians.

In this paper, we propose a method of early detection potential DO students so that the academic staff and faculty also can prevent it immediately. Nevertheless, if the students move to other universities because they have their problems with the services and facilities at their current university, the university can make a plan to deal with the improvements. Good facilities will keep students survive to graduate. While, if economic factors are their reason for move out or DO from, universities can consider providing scholarships, credit or installments payments as a solution.

Many studies have been conducted to detect DO numbers by several researchers. First researched by Andriani [8] related with the use of C4.5 in classifying all students DO. Andriani [8] proposed the attributes were gender, age, origin, middle school, college time, GPA, attendance, parental status, parents' income, scholarships, tuition fees, and work. Andriani [8] does not explain the amount of data used, its characteristics, the purpose of using C4.5, the number of fold cross-validation used, why the evaluation used was AUC, accuracy, precision, recall, etc. Therefore, the results obtained are less convincing to apply to other problems. The second study is a research done by Jaman [9], he was doing student graduation prediction using algorithm C4.5. The attributes used as research are the city of birth, age, gender, province, occupation. A total of 46 data obtained from 2008 records from the Faculty of Computer Science, Universitas Singaperbangsa Karawang, and West Java were used. There were 22 of the total, from the graduating class, while 24 records were those that did not graduate. However, there is no explanation for the reasons for choosing C4.5 and the validation method is not well defined. The characteristics of

978-1-7281-5192-2/19/$31.00 ©2019 IEEE

the data are unclear, thus, the study has not yet finished. In addition, a third study is a research conducted by Kamagi and Hansun [10] which proposed prediction graduation rate of students using C4.5. The data used were alumni students of the 2007-2009 in Informatics Engineering Program at Multimedia Nusantara University, Tangerang, Banten, Indonesia . Data from 2007-2008 range used for training, while the 2009 is used for testing. The attributes proposed by Kamagi and Hansun [10] are IPS (Semester Achievement Index) from the first semester to the sixth semester, gender, secondary school, type of graduation, and the number of semester six credits. Kamagi and Hansun [10] use C4.5 because of their ability to handle discrete and numeric attribute types. The amount of data used is 100, with 60 used for training and 40 for testing. The evaluation proposed to determine the performance of C4.5 is accuracy.

Balanced class distributions, with data types, consist of numerical, polynomial, and binomial. Missing values occur when respondents are reluctant to answer questionnaires because they assume that the question is not too important to fulfill, or even as a result of system failure [11]. Problems can be overcome in a number of ways, including deletion of ambiguous data because it only represents 2% of the dataset, there is an imputation process to resolve missing data because improper data can delete some important information, and missing values can be handled use classification techniques without preprocessing method. One method, for example, is the Decision Tree algorithm [11]. In this paper, the missing values are not preprocessed because the data could be it is characteristic of data from students who have the potential to DO from the university. In this case, Decision Tree not suitable for Unbalanced Class Distribution problems, but can be done for balanced class distribution. Therefore, this paper we propose to solve the most students' problem that leads to potential DO using classification techniques, namely C4.5 to handle missing values and balanced class distribution. C4.5 chosen because it fits with the characteristics of the data used in this study, the data type of numerical and polynomial, in contrast with ID3 (Iterative Dichotomiser 3) which is only suitable for only polynomial data type [12].This research also aims to realize the SDG to shape the quality of education in Indonesia.

## II. Theoretical Framework

### A. Higher Education

One of the mandates given to the Government of the Republic of Indonesia (RI) is to educate its citizens, as stated in the fourth paragraph of the 1945 Republic of Indonesia Law (UUD 45) [13]. To achieve this goal, some of the plans considered for fixing the current problems to improve the quality of students' future masterpiece. The early 1945 constitution stated that "Each citizen has the right to obtain education" and "The government must doing the efforts to organize a national education system regulated by law" [13]. However, in the last 1945 Constitution the fourth amendment was added stating that "Each citizen is obliged to attend elementary education and the government is obliged to finance it" , "The state prioritizes the education budget of at least twenty percent of the state budget (APBN) and APBD (Regional Budget Income and Expenditures) to fulfill the implementation of national education." [13].

There are various ways to realize the Indonesian nation's mandate, including the provision of scholarships for underprivileged and accomplished students. Apart from all this, the rate of DO students still at high rate, especially in the East Java province, Indonesia. According to Higher Education Statistics 2016-2017, the number of dropouts is only reduced by 5,583 or 13% from the previous year [1].

### B. Society 5.0

Technological developments are increasing rapidly and cannot be avoided. Every year, many scientists in various countries continue to develop new concepts in line with the new technological era. Not all Indonesian citizens are ready to face the era of the Industrial Revolution 4.0, even now a new era is emerging, namely, the Society 5.0 proposed by the Japanese government which is more human-centered than the technology-centered in 4.0 Revolution [7].

The following are characteristics of the Industrial Revolution 1.0, 2.0, 3.0 and 4.0 [14]:

1. Industry 1.0: In this era, businesses are processed using mechanical, steam and water power.
2. Industry 2.0: This era has been able to produce products in its business processes, also starting to assemble and use electronic technology.
3. Industry 3.0: In this era computers and automated systems are implemented in business processes.
4. Industry 4.0: Cyber-Physical System (CPS) or physical systems and virtual publications are implemented. This system involves a digital transformation such as Internet Of Things (IOT), Finance Technology (Fintech), Robot, Big Data, Artificial Intelligence (AI), etc. As a pillar of industrial policy [7].

While the era Society 5.0 has the following phases [7] [15]:

1. Society 1.0: People live by hunting and living in harmony with nature.
2. Society 2.0: People who are familiar with the concept of long-term investment with farming.
3. Society 3.0: People living by working in the industrial sector, as private employees, government employees, or entrepreneurs, use industrial technology.
4. Society 4.0: People live by applying information technology to operate their business, thereby increasing various types of jobs such as Database Administrators, Programmers, etc.
5. Society 4.0: People living with maximizing the IoT technology, Fintech, Robot, Big Data, Cloud Computing, Artificial intelligence, etc. In this era there were types of jobs as Youtuber, endorsers, Data Scientists, etc.

### C. Sustainable Development Goals (SDG)

In July 2015, members of the United Nations agreed to the objectives of the Sustainable Development Goals (SDG). The aim is to ensure that people throughout the world can live better without damaging the surrounding environment. Every application of technology must at all times pay attention to the SDG Goals. SDG consists of 17 goals, including [16]:

1. No Poverty
2. Zero Hunger
3. Good Health and Well-Being
4. Quality Education
5. Gender Quality
6. Clean Water and Sanitation
7. Affordable and Clean Energy
8. Decent Work and Economic Growth

9. Industry, Innovation and Infrastructure
10. Reduced Inequalities
11. Sustainable Cities and Communities
12. Responsible Consumption
13. Climate Action
14. Life Below Water
15. Life on Land
16. Peace, Justice and Strong Institutions
17. Partnerships for the Goals

### D. *Missing value*

Missing value is one of the challenges that must be faced in Data Mining, both for supervised learning and unsupervised learning. Missing value occurs due to several reasons, based on reviews Garciarena and Santana [11] which states that, namely 1) Some respondents are not willing to fill out a questionnaire, because the data is considered confidential, not important; with many questionnaire questions, 2) system failure; 3) Data loss during the transfer process, etc.

According to Garciarena and Santana [11] methods can be used to overcome the problem of Missing Values are 1) Data imputation; 2) Deleting the data as long as the number of Missing Values is not significant, for example, 2% of the total dataset, so it does not affect the classification results; 3) Using datasets as they are even though there is a Lost Value by applying the Decision Tree.

### E. *Class Distribution*

A class is said to be a Balanced Class Distribution if it has a distribution between classes with a ratio of 65%: 35% [17]. Not all algorithms are suitable for this, however, one algorithm that is suitable for a Balanced Class Distribution problem is C4.5 which includes families from the Decision Tree.

## III. METHOD

The proposed methods used to detect students dropping out of college are as follows:

1. Data: Data was taken from Academic Information Systems belonging to Private Universities in East Java. Raw data in the form of PDF files and databases that are in the system, so that preprocessing data is done by selecting which attributes are used based on previous research. The data used are data for 2009-2015. The data consisted of 425 students. The data consists of 32 attributes consisting of numerical and nominal data types. Class consists of graduation and dropout. Data is Balanced Class Distribution because it has an inter-class comparison of 44. 94 %: 55.06%. Illustration of data can be seen in Table I.

TABLE I. SUMMARY OF ATTRIBUTES USED

| Attributes | Data Types | Missing Value |
|---|---|---|
| Gender | Binomial | No |
| Lecture Program | Polynomials | No |
| Age when register | Integer | Yes |
| Origin | Polynomials | Yes |
| Year of graduation | Polynomials | Yes |
| Type of last education | Polynomials | Yes |
| Last education status | Polynomials | Yes |
| Major of last education | Polynomials | Yes |
| Marital status | Polynomials | Yes |
| Year of batch | Polynomials | No |
| History of leave of absence | Polynomials | No |
| Number of courses | Integer | No |
| Credit unit semester 1 | Integer | Yes |
| Credit unit semester 2 | Integer | Yes |
| Credit unit semester 3 | Integer | Yes |
| Credit unit semester 4 | Integer | Yes |
| Credit unit semester 5 | Integer | Yes |
| Credit unit semester 6 | Integer | Yes |
| Credit unit semester 7 | Integer | Yes |
| Credit unit semester 8 | Integer | Yes |
| Credit unit semester 9 | Integer | Yes |
| Credit unit semester 10 | Integer | Yes |
| GPA semester 1 | Real | Yes |
| GPA semester 2 | Real | Yes |
| GPA semester 3 | Real | Yes |
| GPA semester 4 | Real | Yes |
| GPA semester 5 | Real | Yes |
| GPA semester 6 | Real | Yes |
| GPA semester 7 | Real | Yes |
| GPA semester 8 | Real | Yes |
| GPA semester 9 | Real | Yes |
| GPA semester 10 | Real | Yes |

2. Classification using C4.5 with 4 criteria. Criteria is one of the parameters of the family of decision tree algorithms including C4.5 which consists of Gain ratio, information gain, gini index, and accuracy. This criterion serves to determine the attributes/features/ variables that are right for breaking the node [18, 19, 20, 21].
3. Validation uses 10-fold cross validation with three sampling methods (Linear, Shuffled, and Stratified) [22].
4. Evaluation using accuracy, precision, recall, and AUC [23]. A classification algorithm is said to have good performance if it has an AUC value close to 1, illustrating the meaning of AUC shown in Table II.

TABLE II. AUC VALUE, MEANING AND SYMBOL [24]

| AUC | Meaning |
|---|---|
| 0,9 - 1 | Excellent classification |
| 0,8 – 0,9 | Good classification |
| 0,7 – 0,8 | Fair classification |
| 0,6 – 0,7 | Poor classification |
| < 0,6 | Failure |

## IV. RESULTS AND DISCUSSION

Based on the test results, the C4.5 algorithm has successfully detected all-student College that had the potential of DO with the data that has characteristics Missing Values, shown in Fig. 1 and Fig. 2.

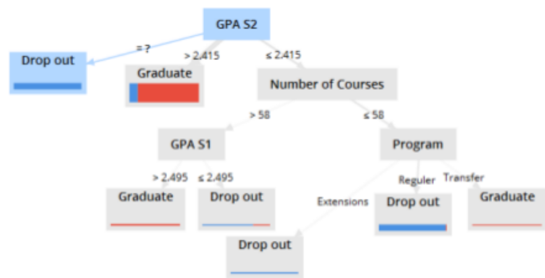


Fig. 1. Results of C4.5 in the data of the first year of lecture.

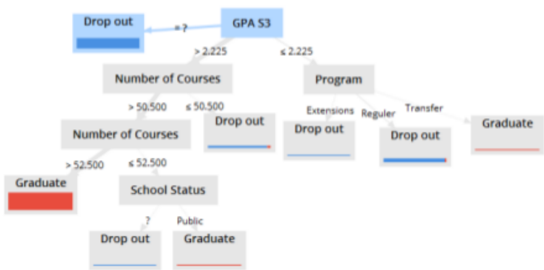


Fig. 2. Results of C4.5 in the data for the third year of lecture

The test is conducted 60 times, the first one focuses on the C4.5 parameter criteria method, the second is the Sampling method, while the third focuses on the study year. The first test is based on parameter C4.5 using four criteria, namely Gain ratio, information gain, gini index, accuracy. It aims to determine which criteria are the most optimal for the data used in this study. The results showed that the most optimal criteria were dominated by using accuracy, while the use of the Gain ratio criteria had the worst results, consider Fig. 3.

The second test is based on sampling methods, including Linear (L), Shuffled (Sh), and Stratified (St). The results showed that the Linear method had the worst performance because Linear produced random data whose class distribution between classes in the training data as well as the test data class was not balanced, while C4.5 was not suitable for unbalanced Class Distribution, as illustrated in Table I and Fig. 4. While the method that has optimal performance is Stratified, because randomly generated between training data and test data has a balanced distribution class, while C4.5 is suitable for balanced class distribution.

The third test is based on the academic year consisting of the first, second, third, fourth, and fifth years. Detection of students with potential DO has optimal results when using data in the third year during their studies in the case of Balanced Class Distribution as indicated by applying stratified sampling, consider Table III numbers in bold. While the application of data based on the first year study resulted in very poor average performance, especially in the Unbalanced Class Distribution data indicated by applying Linear Sampling, consider Table III with the numbers outlined in the section.

In Fig. 1 , all regular and extension students in second semester have GPA ≤ 2,415 and has taken ≤ 58 courses become 'Drop Out', while transfer students will 'Graduated' . This happens the number of subjects recorded in this study is the total number of courses taken by students who have DO and students who have graduated, not the total number of eyes lecture every semester. So that even though it was still the first year that normally only took 6 courses, and 16 of them ≤ 58 is considered DO, while for transfer students, sometimes they only go to one semester to continue D3 to S1, which in 1 year is subject ≤ 58, and on average they graduate. So the pattern obtained by C4.5 is the same, not common. In the next study the number of each subject is expected the number of courses taken each semester, not cumulatively during the study.

TABLE III. C4.5 Test Results

| **Year of Lecture** | ***Sample*** | **Accuracy (%)** | **Precision (%)** | ***Recall* (%)** | **AUC** |
|---|---|---|---|---|---|
| 1st year | *Linear* | 80,95 | 58,29 | 80,88 | 0,09 |
| | *Shuffled* | 86,82 | 87,50 | 89,05 | 0,87 |
| | *Stratified* | 86,30 | 87,49 | 87,95 | 0,87 |
| 2nd year | *Linear* | 90,89 | 59,32 | 91,78 | 0,14 |
| | *Shuffled* | 91,26 | 92,66 | 90,94 | 0,92 |
| | *Stratified* | 92,39 | 93,23 | 92,13 | 0,93 |
| **3rd year** | *Linear* | 96,78 | 88,06 | 96,69 | 0,19 |
| | *Shuffled* | **96,94** | 97,44 | 96,85 | 0,95 |
| | ***Stratified*** | 96,87 | **98,75** | **97,84** | **0,98** |
| 4th year | *Linear* | 95,85 | 78,70 | 95,19 | 0,19 |
| | *Shuffled* | **96,94** | 97,58 | 96,78 | 0,95 |
| | *Stratified* | 96,87 | 97,61 | 96,89 | 0,93 |
| 5th year | *Linear* | 95,50 | 78,34 | 95,30 | 0,23 |
| | *Shuffled* | 96,06 | 96,10 | 96,66 | 0,95 |
| | *Stratified* | 96,70 | 97,71 | 96,35 | 0,95 |

Based on the results of the trial it can be proven that C4.5 is suitable for Balanced Class Distribution problems, both data are numerical or polynomial. Although in this case it has the characteristics of Balanced Class Distribution, but the selection of random sampling techniques is very decisive, which is the best stratified in third year students in detecting students dropping out of college , whereas Linear has the worst performance in the first year of lecture . In the case of this study can be handled by a missing value C4.5 that are the hallmarks of a derivative of Decision Tree and cases of missing value in the application of this C4.5 a feature or pattern of students who potentially drop out as presented in Fig. 1 and Fig. 2.

## V. Conclusion

Decision Tree Algorithm is applied to classification by means of data that has a missing value and without an imputation process. In addition, the Decision Tree algorithm is good for balanced class distribution. C4.5 in this study has a very good performance when using data in the third year of college with elections stratified random sampling, with the worst performance in the first year of college with linear selection random sampling. This research can be used as the first step in welcoming the Society 5.0 era which maximizes the data that has been collected in the databases of each university. This research is intended to support one of the 17 SGD objectives set by the United Nations in the field of improving the quality of education and to realize the mandate of the 1945 Constitution.

Because the problem of missing values has the potential to predict the majority of students who drop out, further research can analyze the data with a focus on outlier problems . Outlier can be determined by using another interpolation method to get some samples that do not fit between populations. Further studies expected to present the number of subject / courses taken each semester, not cumulatively during the study from the beginning to the end.

## Acknowledgment

Our appreciation and gratitude goes to the Directorate of Research and Community Service, Director General of Research Strengthening and Development, Ministry of Research, Technology and Higher Education, for funding this research Cooperation Between Universities in the second year, with the 2019 title "Educational Data Mining

Based on Classification for Analyzing Students with the Potentially to Drop Out from College in Case of Imbalanced Class Distribution".

## REFERENCES


[1] P. I. D. Kemenristekdikti, "Statistik Pendidikan Tinggi Tahun 2017," Kemenristekdikti, Jakarta, 2017.

[2] W. Purba, S. Tamba and J. Saragih, "The effect of mining data k-means clustering toward students profile model drop out potential," *Journal of Physics: Conference Series,* vol. 1007, no. 1, p. 012049, 2018.

[3] J. R. Casanova, A. Cervero, J. C. Núñez, L. S. Almeida and A. Bernardo, "Factors that determine the persistence and dropout of university students," *Psicothema,* vol. 30, no. 4, pp. 408-414, 2018.

[4] M. R. K. Milon, M. R. Hossain and M. R. Alam, "Factors Influencing on Dropouts at Undergraduate Level in Private Universities of Bangladesh: A Case Study," *Australasian Journal of Business, Social Science and Information Technology,* vol. 4, no. 4, pp. 177-188, 2018.

[5] D. Kim and S. Kim, "Sustainable Education: Analyzing the Determinants of University Student Dropout by Nonlinear Panel Data Models," *Sustainability,* vol. 10, no. 4, 2018.

[6] M. R. Larsen, H. B. Sommersel and M. S. Larsen, Evidence on Dropout Phenomena at Universities, Copenhagen, Nordvest: Aarhus University, 2013.

[7] M. Fukuyama, "Society 5.0: Aiming for a New Human-Centered Society," *Japan SPOTLIGHT,* pp. 47-50, 2018.

[8] A. Andriani, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," in *Seminar Nasional Matematika 2012*, Surakarta, 2012.

[9] J. H. Jaman, "Prediksi Kelulusan Mahasiswa Dengan Metode Algoritma C4.5," *SYNTAX,* vol. 2, no. 2, pp. 39-43, 2013.

[10] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *ULTIMATICS,* vol. 6, no. 1, pp. 15-20, 2014.

[11] R. S. U. Garciarena, "An extensive analysis of the interaction between missing data types, imputation methods, and supervised classifiers," *Expert Systems with Applications,* vol. 89, pp. 52-65, 2017.

[12] B. Hssina, A. Merbouha, H. Ezzikouri and M. Erritali, "A comparative study of decision tree ID3 and C4. 5," *International Journal of Advanced Computer Science and Applications,* vol. 4, no. 2, pp. 13-19, 2014.

[13] T. R. BIP, "Undang-Undang Dasar Negara Republik Indonesia Tahun 1945," Bhuana Ilmu Populer, Jakarta, 2017.

[14] D. Vuksanovic, J. Ugarak and D. Korčok, "Industry 4.0: The future concepts and new visions of factory of the future development," in *Sinteza 2016 - International Scientific Conference on ICT and E-Business Related Research*, Belgrade, 2016.

[15] T. G. o. Japan, "Realizing Society 5.0," 11 September 2018. [Online]. Available: https://www.japan.go.jp/abenomics/_userdata/abenomics/pdf/society_5.0.

[16] M. LAC, The World We Want: A Guide to the Goals for Children and Young People, Panama: Unicef, 2015.

[17] H. Li and J. Sun, "Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples–Evidence from the Chinese hotel industry," *Tourism Management,* vol. 33, no. 3, pp. 622-634, 2012.

[18] H. M. Asmara, "Analisa Perbandingan Hasil Pohon Keputusan dengan Gain Ratio, Information Gain, dan Gini Index pada Pemasaran Produk Herbal di CV. Al-Ghuroba," Universitas Muhammadiyah Surakarta, Surakarta, 2016.

[19] R. C. Barros, A. C. d. Carvalho and A. A. Freitas, Automatic Design of Decision-Tree Induction Algorithms, Springer, 2015.

[20] J. Han, M. Kamber and J. Pei, Data mining concepts and techniques, Waltham: Morgan Kaufmann, 2011.

[21] D. T. Larose and C. D. Larose, Discovering knowledge in data: An introduction to data mining, 2nd ed., New Jersey: John Wiley & Sons, 2014.

[22] S. Mutrofin, A. Mu'alifa, R. V. H. Ginardi and C. Fatichah, "Solution of class imbalance of k-nearest neighbor for data of new student admission selection," *International Journal Of Artificial Intelegence Research,* vol. 3, no. 2, pp. 47-55, 2019.

[23] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran and S. Venkatraman, "Robust Intelligent Malware Detection Using Deep Learning," *IEEE Access,* vol. 7, pp. 46717- 46717, 2019.

[24] R. S. Wahono, N. S. Herman and S. Ahmad, "A comparison framework of classification models for software defect prediction," *Advanced Science Letters,* vol. 20, no. 10-12, pp. 1945-1950, 2014.

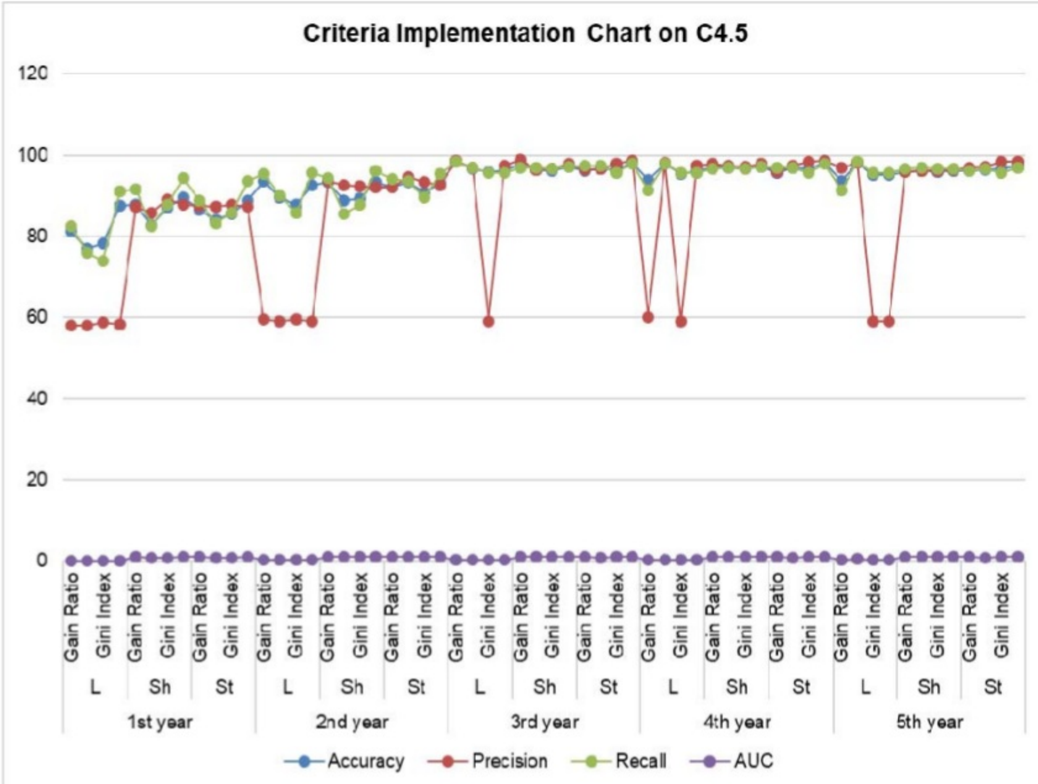


Fig. 3 C4.5 results based on selection criteria

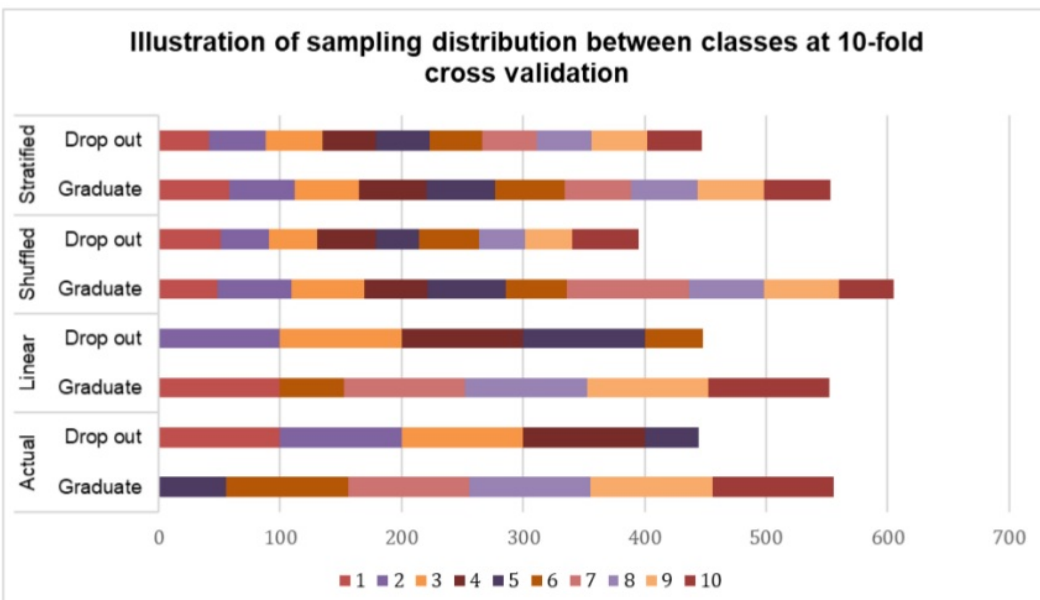


Fig. 4 Inter-class sampling distribution

# Detection of Potentially Students Drop Out of College in Case of Missing Value Using C4.5

ORIGINALITY REPORT

| 10% | 7% | 7% | 4% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.childreach.org.uk<br>Internet Source | 1% |
| 2 | ijair.id<br>Internet Source | 1% |
| 3 | eprints.unipdu.ac.id<br>Internet Source | 1% |
| 4 | Submitted to Universitas Dian Nuswantoro<br>Student Paper | 1% |
| 5 | eudl.eu<br>Internet Source | 1% |
| 6 | www.iges.or.jp<br>Internet Source | 1% |
| 7 | Muhamad, Abdul, Muhammad Ramli, Manan Ceylon, and Abdullah Sindring. "The Apparatus Performance of Public Administration in Education and Culture Field at Bone Bolango Regency of Gorontalo Province", Mediterranean Journal of Social Sciences, 2016. | 1% |

| | | |
|---|---|---|
| | Publication | |
| 8 | worldwidescience.org<br>Internet Source | <1% |
| 9 | www.adb.org<br>Internet Source | <1% |
| 10 | Submitted to University of Melbourne<br>Student Paper | <1% |
| 11 | Yuita Arum Sari, Fitri Utaminingrum, Sigit Adinugroho, Ratih Kartika Dewi et al. "Indonesian Traditional Food Image Identification using Random Forest Classifier based on Color and Texture Features", 2019 International Conference on Sustainable Information Engineering and Technology (SIET), 2019<br>Publication | <1% |
| 12 | jurnal.kdi.or.id<br>Internet Source | <1% |
| 13 | repository.uinsu.ac.id<br>Internet Source | <1% |
| 14 | Sugandi Yahdin, Anita Desiani, Ali Amran, Desty Rodiah, Solehan. "Pattern recognation for study period of student in Mathematics Department with C4.5 algorithm data mining technique at the Faculty of Mathematics and Natural Science Universitas Sriwijaya", Journal | <1% |

| | | |
|---|---|---|
| | of Physics: Conference Series, 2019<br>Publication | |
| 15 | Siti Mutrofin, Maisarah Maisarah, Slamet Widodo, Raden Venantius Hari Ginardi, Chastine Fatichah. "Application of a combination between Principal Component Analysis and Logistic Regression Based on Support Vector Machine on Educational Data Mining with Overlapping Data Problem", IOP Conference Series: Materials Science and Engineering, 2020<br>Publication | <1% |
| 16 | docplayer.info<br>Internet Source | <1% |
| 17 | budi.rahardjo.id<br>Internet Source | <1% |
| 18 | www.jurnal.ar-raniry.ac.id<br>Internet Source | <1% |
| 19 | "1st Annual Conference of Midwifery", Walter de Gruyter GmbH, 2020<br>Publication | <1% |
| 20 | www.amrita.edu<br>Internet Source | <1% |
| 21 | publicadministration.un.org<br>Internet Source | <1% |

| | | |
|---|---|---|
| 22 | Kapil Narula. "Ensuring Sustainable Oceans", Maritime Affairs: Journal of the National Maritime Foundation of India, 2015<br>Publication | <1% |
| 23 | "Finalproceedings", 2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), 2019<br>Publication | <1% |
| 24 | Siti Mutrofin, Ainul Mu'alif, Raden Venantius Hari Ginardi, Chastine Fatichah. "Solution of class imbalance of k-nearest neighbor for data of new student admission selection", International Journal of Artificial Intelligence Research, 2019<br>Publication | <1% |

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | On | | |

# Detection of Potentially Students Drop Out of College in Case of Missing Value Using C4.5

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

**Instructor**

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6